REMARKS

In the Office Action, the Examiner rejected Claims 1-15, 19-23 and 27-29, which were all of the then pending claims, under 35 U.S.C. 103 as being unpatentable over U.S. Patent 6,405,161 (Goldsmith) in view of U.S. Patent 7,103,536 (Kanno).

Independent Claims 1, 13 and 21 are being amended to better define the subject matters of these claims. In particular, these claims are being amended to describe in more detail the way in which Patricia trees are formed and used to identify sets of candidate prefixes and candidate suffixes Claim 12 is being cancelled to reduce the number of issues in this application, and new Claim 30, which is dependent from Claim 13, is being added to describe a preferred feature of the invention.

For the reasons advanced below, Claims 1-11, 13-15, 19-23 and 27-30 patentably distinguish over the prior art and are allowable. The Examiner is thus asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15, 19-23 and 27-29 under 35 U.S.C. 103, and to allow these claims and new Claim 30.

Generally, Claims 1-15, 19-23 and 27-29 patentably distinguish over the prior art because the prior art does not show or suggest forming and using Patricia trees, as described in independent Claims 1, 13 and 21, to identify sets of candidate prefixes and candidate suffixes,

Applicants' invention, generally, relates to automatically collecting affixes of a language from one or more documents. As discussed in detail in the present application, knowledge of affixes is important for analyzing existing words and for producing new words. It is very difficult and time extensive to acquire a complete list of affixes of a language by hand, and a number of procedures have been tried to use a computer or some other automated process to identify affixes.

The previous approaches tend to parse words into pieces, either a prefix and a stem, or a stem and a suffix. Also, the prior art approaches tend to limit the length of an affix to reduce the size of the search space.

The prior art systems have a number of disadvantages and limitations. For example, they may fail to discover both prefixes and suffixes at the same time, and they may not be able to discover nested affixes. In addition, because of the limitations on length, the prior art systems may not find many affixes that appear in technical documents. Further, many of the previous approaches fail to find affixes containing non-alphabet characters such as digits and hyphens.

The present invention effectively addresses these prior art limitations. Generally, the present invention provides an unsupervised, knowledge-free procedure for automatically discovering prefixes and suffixes from text. The present invention integrates prefix and suffix discovery in such a way that uses knowledge about prefixes to find suffixes and uses knowledge about suffixes to find prefixes.

More specifically, in one embodiment, the invention provides a computer system for analyzing text in one or more documents, and comprising one or more system interfaces, and an affix process that determines one or more affixes of one or more words in one or more of the documents and provides the affixes to the system interface. This affix determining process comprises obtaining a collection of words, and representing all of the words in the collection as Patricia trees to show visually morphological structures of the words.

To form these Patricia trees, the words in the collection are first used to construct first and second tries. Each of these tries has a multitude of paths and a multitude of nodes, and these tries are then compresses by compressing all unary paths in the tries to form the prefix Patricia tree and the suffix Patricia tree. In this way, words are added into the Prefix tree, and the Prefix

I:\IBM\105\16751\amend\16751.am4.doc

tree is used to identify a set of candidate prefixes. Also, in the formation of the suffix Patricia tree, the words in the prefix Patricia tree are reversed, and the reversed words are added into the suffix Patricia tree. This suffix Patricia tree is then used to identify a set of candidate suffixes.

These sets of candidate prefixes and suffixes are then refined, using knowledge previously discovered, to identify actual prefixes and suffixes. In particular, knowledge of prefixes previously identified in the refining is used to further refine the set of candidate suffixes, and knowledge of suffixes previously identified in the refining is used to further refine the set of candidate prefixes.

The references of record do not disclose or render obvious forming and using Patricia trees in the above-described manner to identify sets of candidate prefixes and sets of candidate suffixes.

For instance, Goldsmith discloses an automated, morphological analysis of a natural language for determining prefixes, suffixes and stems. This analysis has three major components: determining the correct morphological split for individual words, establishing accurate categories of stems based on the range of suffixes that they accept, and identifying allomorphs of the same stem.

In the procedure described in Goldsmith, word stems and suffixes are identified and combined as signatures, and then prefixes are identified in the list of stems and suffixes. As the Examiner noted in the Office Action, column 5, lines 51-61, of Goldsmith discloses that prefixes are identified by removing common sequences of one or more letters from the above-mentioned signatures.

As the Examiner has also recognized, there are a number of important features of Applicants' invention that are not shown in or rendered obvious by Goldsmith. In particular,

Goldsmith does not disclose the use of Patricia trees to identify candidate sets of prefixes and suffixes. In order to remedy this deficiency of Goldsmith as a reference, the Examiner relies on Kanno.

Kanno describes a process for identifying a suffix, and the Examiner cited this reference for the disclosure of reversing a word and adding it to a Patricia tree for suffix matching. Applicants' process involves much more than reversing words and adding those reversed words to a Patricia tree.

With the present invention, the Patricia trees are formed by initially forming first and second tries and then compressing these tries to form the prefix Patricia tree and the suffix Patricia tree. In particular, in Applicants' invention, each of the tries has a multitude of paths and a multitude of nodes, and the tries are compressed by compressing all unary paths in the tries to form the prefix and suffix Patricia trees. In this process, words are added into the prefix Patricia tree; and, in the formation of the suffix Patricia tree, the words in the prefix Patricia tree are reversed, and the reversed words are added into the suffix Patricia tree. The Prefix tree is used to identify a set of candidate prefixes, and the suffix Patricia tree is used to identify a set of candidate suffixes.

This process of the present invention is of utility because, as discussed in the application, it enables the potential candidates of prefixes and suffixes to be easily identified.

The other references of record have been reviewed, and these other references, whether considered individually or in combination, also do not disclose or suggest the refinement process of the present invention.

For example, U.S. patent application publication no. 2003/0105638 (Taira), which was cited in the previous Office Action, discloses a method and system for generating structured

medical reports from natural language reports. In the procedure described in Taira, statistical natural language processing is used to analyze individual words and combinations of words to generate more accurate translations. Paragraph (74) of Taira refers to a recursive algorithm, in which the results of one iteration are sent to the next stage of processing, and based on the results of this stage, may trigger a refinement of a word's classification. There is no disclosure, however, in Tiara of forming and using Patricia trees as they are formed and used in the present invention.
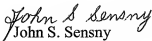
Independent Claims 1, 13 and 21 are being amended to emphasize the above-described aspect of the present invention. In particular, each of Claims 1, 13 and 21, as amended herein, describes the feature of representing all of the words in a collection as Patricia trees to show visually morphological structures of the words, including using the words to construct first and second tries, and compressing the first and second tries by compressing all unary paths on the tries to form a prefix Patricia tree and a suffix Patricia tree. Claims 1, 13 and 21 positively set forth the additional limitation that the prefix and suffix Patricia trees are used, respectively, to identify a set of candidate prefixes and a set of candidate suffixes.

In view of the above-discussed differences between Claims 1, 13 and 21 and the prior art, and because of the advantages associated with those differences, Claims 1, 13 and 21 patentably distinguish over the prior art and are allowable. Claims 2-11 are dependent from Claim 1 and are allowable therewith. Also, Claims 14, 15, 19, 20, 29 and 30 are dependent from Claim 13 and are allowable therewith; and Claims 22, 23, 27 and 28 are dependent from, and are allowable with, Claim 21.

The Examiner is, accordingly, asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15, 19-23 and 27-29 under 35 U.S.C. 103, and to allow these claims and new

Claim 30. If the Examiner believes that a telephone conference with Applicants' Attorneys would be advantageous to the disposition of this case, the Examiner is asked to telephone the undersigned.

Respectfully Submitted,

John S Sensny
John S. Sensny
Registration No. 28,757
Attorney for Applicant

Scully, Scott, Murphy & Presser, P.C.
400 Garden City Plaza, Suite 300
Garden City, New York 11530
(516) 742-4343

JSS:gc